

# Voorstel: Objectidentificatie binnen stelsel van basisregistraties

(versie 2014-02-28)

---

## Inleiding

*Korte beschrijving van het onderwerp.*

Registraties zijn een digitale representatie van werkelijkheid. De representatie vindt plaats middels digitale objecten met kenmerken die dingen met eigenschappen in de werkelijkheid representeren. Een principale eigenschap van een registratie of register is dat digitale objecten uniek identificeerbaar moeten zijn. Uniek binnen een registratie maar als de gegevens extern gepubliceerd worden ook uniek over registraties heen. In feite uniek identificeerbaar binnen het publicatie of gebruikdomein. Nu is voor basisregistraties dat gebruikdomein onbepaald en in die zin de hele wereld. Unicité is daarom een onderwerp dat over registraties heen geadresseerd moet worden. Dit onderwerp behandelt een systematiek voor unieke objectidentificatie. Ter verduidelijking, het gaat hierbij niet om de identificatie van 'dingen' in de werkelijkheid, maar van hun digitale representatie als objecten.

## Analyse

*Korte beschrijving van de huidige situatie.*

Identificatie van objecten is een onderwerp dat in alle basisregistraties een rol speelt en in de specificatie van een registratie dient dit beschreven te zijn.

Voor alle standaarden is het nodig dat de objectidentificatie uniek is, anders kunnen er in het berichtenverkeer geen gegarandeerd valide referenties worden gemaakt. De verantwoordelijkheid hiervoor ligt nu bij de registraties. Elke registratiehouder heeft hiervoor een eigen ontwikkelde systematiek.

Het ontbreken van een gezamenlijke systematiek betekent dat er geen garantie is voor unieke objectreferenties tussen objecten in verschillende registraties. Met andere woorden dat er geen garantie is voor correcte verwijzingen tussen registraties. Dat is een zwakke schakel in het *stelsel*.

## Voorstel: Nationale Strategie voor unieke objectidentificatie.

NEN 3610 heeft een systematiek voor het voorzien in unieke identifiers binnen het NEN 3610 domein middels het creëren van een namespaces. Registraties worden gekoppeld aan unieke namespaces. Binnen een namespace is de registratie zelf verantwoordelijk voor het uitdelen van unieke identifiers (uniek binnen de registratie). Er ontstaat op die manier een getrapte verantwoordelijkheid voor unieke identifiers. Dat kan echter alleen maar werken indien er een nationaal register van namespaces is. Het ontbreekt op dit moment aan een dergelijke afspraak.

Over dit principe is verder nagedacht in het programma Linked Open Data. Het resultaat daarvan is de URI strategie voor linked data. Kern hiervan is een URI patroon dat verwijst naar een domein, een type identifier, een concept en de uiteindelijk referentie. In feite een namespace – of domeinnaamregister uitgebreid met een concept- en object register. Het ligt voor de hand om eenzelfde strategie ook toe te passen voor data die niet middels het linked data concept gepubliceerd zijn.

Dit voorstel presenteert de Nationale URI strategie als best practice voor unieke objectidentificatie.

Nationale URI-strategie:

[http://www.pilod.nl/wiki/Boek/URI-strategie#De\\_URI-strategie](http://www.pilod.nl/wiki/Boek/URI-strategie#De_URI-strategie)

## De URI-strategie

### Toepassingsgebied:

De URI-strategie is met name bedoeld voor data waarmee objecten of concepten worden gedefinieerd, waar andere toepassingen naar kunnen verwijzen. Dit betekent dat het data betreft die extern zijn gepubliceerd.

De termen 'data' en 'concepten' verwijst naar de algemene toepassing van een URI-strategie waarin de volgende drie categorieën concepten een rol spelen:

1. Begrippen in een conceptueel model
2. Referentieobjecten
3. 'gewone' data

Het aspect objectidentificatie heeft alleen betrekking op het laatste concept 'gewone data'. Voor de concepten 'begrippen en referentieobjecten' kan dezelfde methodiek toegepast worden maar dat is geen onderwerp van deze afspraak.

**URI-patroon:** (bron: [http://www.pilod.nl/wiki/Boek/URI-strategie#De\\_URI-strategie](http://www.pilod.nl/wiki/Boek/URI-strategie#De_URI-strategie))

In feite kan volstaan worden met verwijzing naar bovenstaande url. Voor de gedachtebepaling rond het onderwerp zijn een aantal onderdelen hier over genomen.

De objectidentificatie voor externe publicatie van objecten volgt het volgende patroon.

`http://{domain}/{type}/{concept}/{reference}`

Dit patroon resulteert in een getrapte opbouw van uniciteit. De onderdelen daarvan zijn:

#### **{domain}**

Het {domain} deel bevat het internet domein en eventueel een pad binnen dat domein

`{domain} = {internet domain}/{path}.`

Het {domain} dient twee doelen. Ten eerste is het een belangrijk instrument om unieke identificaties te verkrijgen: twee objecten, die beheerd worden in twee verschillende databases, kunnen toevallig dezelfde identificatie krijgen (bijvoorbeeld een kadastraal perceel met id 010101 en een rechtspersoon met id 010101). Als nu zowel het Kadaster als het Nieuw HandelsRegister (NHR) besluit om deze objecten als data te publiceren, worden er toch twee unieke URIs gevormd: de een begint bijvoorbeeld met `http://brk.nl/` en de ander met `http://nhr.nl/`. Ten tweede zorgt een goedgekozen domein voor herkenbaarheid en vertrouwen. Kadastrale percelen met een URI als `http://data.brk.nl/perceel/010101` hebben een betrouwbaarder uitstraling dan bijvoorbeeld `http://data.vindhethier.eu/perceel/010101`.

Het {path} kan worden gebruikt als binnen een register verschillende verzamelingen objecten leven, waarbij dubbele id's kunnen voorkomen. Het {path} kan dan gebruikt worden om extra namespaces te creëren.

#### **{type}**

Het {type} geeft aan om wat voor soort URI het gaat. Voor een objectidentificatie is dat:

'id' = identifier van een object (individual/instance) in een register.

### **{concept}**

Het {concept} geeft de menselijke lezer een indicatie van het soort concept dat de URI identificeert. Het {concept} is belangrijk om twee redenen. Ten eerste kan het een uitkomst bieden als objecten binnen de registratie geen unieke identifiers hebben, maar wel uniek zijn per soort object. Bijvoorbeeld gemeente Utrecht en provincie Utrecht. Ten tweede, en dit is belangrijker, levert het een begrijpelijker URI op. Een menselijke lezer kan vermoeden dat <http://bagregister.nl/id/pand/01010101> de URI van een pand uit de BAG is.

Een mogelijk nadeel van het opnemen van {concept} in de URI is dat hiermee betekenis in de URI wordt opgenomen, terwijl betekenisloze IDs over het algemeen eenvoudiger persistent te maken zijn.

#### Aanbevelingen voor {concept}

1. {concept} heeft geen inhoudelijke betekenis.  
Het is onverstandig om {concept} enige betekenis toe te kennen voor de machine. URIs zijn in technische zin transparant. Het is dus niet zo dat het {concept} per se de klasse is waartoe een object behoort. Het helpt alleen de menselijke lezer, bijvoorbeeld de beheerder van een semantisch model, om de URIs te herkennen.
2. Denk ook bij het kiezen van {concept} aan persistentie.  
Als het in een registratie denkbaar is dat objecttypen (klassen) van naam veranderen, maar dan nog wel dezelfde klasse vertegenwoordigen, is het niet verstandig dit onderdeel in de URI op te nemen. Neem in dat geval een hogere klasse op. Volgens sommigen betekent het veranderen van het type van een instance per definitie dat er niet langer sprake is van dezelfde instance, maar van een andere instance, van het andere type. Voorbeeld: stel dat het Centraal Orgaan opvang Asielzoekers (COA) wordt omgevormd van zelfstandig bestuursorgaan (zbo) naar agentschap. En dat we als URI van het COA zouden kiezen voor: {domein}/id/zbo/coa. Dan wordt dat na de omvorming {domein}/id/agentschap/coa. Zouden we kiezen voor {domein}/id/organisatie/coa dan hoeven we de URI niet aan te passen, maar kunnen we op basis van de uri ook geen onderscheid meer maken tussen de COA als ZBO en de COA als agentschap.

### **{reference}**

De {reference} is de identificerende naam of code van het individuele object. Wat betreft {reference} geeft de URI strategie veel vrijheid, aangezien de eisen in verschillende toepassingen sterk uiteen kunnen lopen. Een {reference} kan zijn: een identificerend nummer, een alfanumerieke code, een woord of naam, etc. Elk register heeft wel een manier om de individuele objecten in de verzameling uniek aan te duiden. Deze unieke aanduiding kan worden opgenomen in de {reference}.

#### Aanbevelingen voor {reference}

1. Namen of nummers?  
Er is vaak discussie over het gebruik van 'betekenisloze' identifiers versus 'betekenisvolle' identifiers. Zolang computers geen bewustzijn hebben is elke URI voor de machine een betekenisloze string. Voor mensen kan ook een betekenisloze string betekenis krijgen (020 wordt veel gebruikt door mensen die het label 'Amsterdam' of 'Ajax' niet willen uitspreken, 013 (Tilburgs poppodium), 9292 (OV-informatie), nummer 14 (Johan Cruijff).

Namen of nummers, voor beiden is wat te zeggen. Nummers heeft als voordeel dat het nauwkeuriger lijkt en er geen homoniemen voor kunnen komen. Maar je verliest herkenbaarheid en hanteerbaarheid voor mensen, zonder steeds de labels bij de hand te hebben.

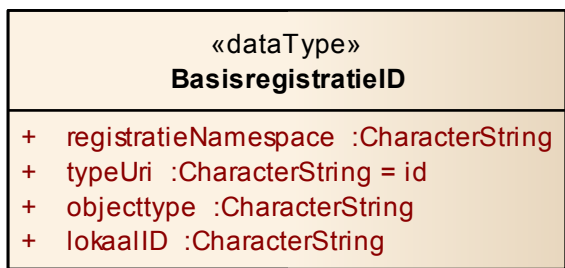
- In de praktijk zijn de URIs voor de concepten in vrijwel alle semantische standaarden betekenisvol en bevatten zij doorgaans het volledige label (naam) waarmee de term voor de mens wordt aangeduid (meestal als CamelCase geschreven zodat er geen spaties in voorkomen).
- Bij grote aantallen objecten wordt het ondoenlijk om voor elk object een herkenbare unieke naam te bedenken. We gaan dan - vrijwel vanzelf - nummers.

- Tussen deze twee uitersten zit een grijs gebied. Voor kleine, stabiele sets met objecten (bijvoorbeeld provincies) is het voordelig om de hele naam in de URI op te nemen, bij iets grotere sets, met meer mutaties, komen vaak lange namen voor die de URI onhandelbaar maken. Het kan dan een oplossing zijn om afkortingen in de URI te gebruiken.
- 2. Vermijd vreemde tekens in een URI.  
Het wordt aangeraden om zich te beperken tot lowercase letters, cijfers, en eventueel koppeltekens als scheidingstekens.

## Implementatie

### In dataspecificaties:

De URI-strategie is voor externe publicatie van objecten. Registraties dienen in de specificaties op te nemen wat de onderdelen zijn waar de externe identifier uit is opgebouwd. In feite zijn dat de onderdelen van het uri-patroon. Een voorstel voor de implementatie is creatie van een samengesteld datatype met de naam 'BasisregistratieID'. De daarin voorkomende attributen komen overeen met de onderdelen van het URI-patroon maar hebben een andere benaming. Die andere benaming is gekozen omdat die beter aansluit bij de functie als unieke identifier van objecten.



- |                                 |                       |
|---------------------------------|-----------------------|
| - registratieNamespace = domein | bijvoorbeeld 'bag.nl' |
| - typeUri = type                | waarde is altijd 'id' |
| - objecttype = concept          | bijvoorbeeld 'pand'   |
| - lokaalID = reference          | bijvoorbeeld '233344' |

### In XML berichtenverkeer:

In berichten wordt de identifier geïmplementeerd in het XML datatype xsd:id. De codering daarvan is conform de volgende regel:

xsd:id= "http://{registratieNamespace}/{typeUri}/{objecttype}/{lokaalID}"

Voorbeeld:

xsd:id= <http://bag.nl/id/pand/233344>

Een uitbreiding hierop is het toevoegen van een namespace voor de gezamenlijke basisregistraties.

Hiermee wordt een gezamenlijke verantwoordelijkheid aangegeven. Het voorbeeld hierboven wordt dan:

xsd:id= <http://basisregistratie.nl/bag/id/pand/233344>

## Aanpak

*Hoe komt de standaard voor dit onderwerp tot stand?*

*Wat moet er gebeuren? Wie gaat het doen en hoeveel inspanning kost het?*

Bij de URI-Strategie is een aantal open issues opgenomen. Eén daarvan is de volgende:

Mate waarin de strategie geformaliseerd dient te worden:

Het ultieme doel is een situatie waarin elk object maar 1 URI heeft. Die URI wordt gemunt door de authentieke registratie van dat object. We zouden formele criteria waaraan een register moet voldoen als volgt kunnen formuleren:

Een register:

- is een Standaard of Authentieke Registratie
- heeft een formele basis voor authenticiteit:
- de Standaard komt voor op de 'pas toe of leg uit'-lijst met open standaarden van het Forum Standaardisatie
- de Authentieke Registratie wordt als zodanig aangeduid in Nederlandse of EU-regelgeving
- heeft een registerhouder
- heeft een eigen internetdomein (namespace)
- heeft een eigen URI-patroon dat voldoet aan de nationale URI-strategie

Dit geldt ook voor de URI-strategie zelf: Kan de URI-strategie als standaard worden gedefinieerd? En zo ja, is het dan voldoende om deze standaard op de 'pas toe of leg uit'-lijst van het Forum Standaardisatie te plaatsen of zijn er uitgebreidere maatregelen gewenst, misschien zelfs wetgeving?